Computer Vision

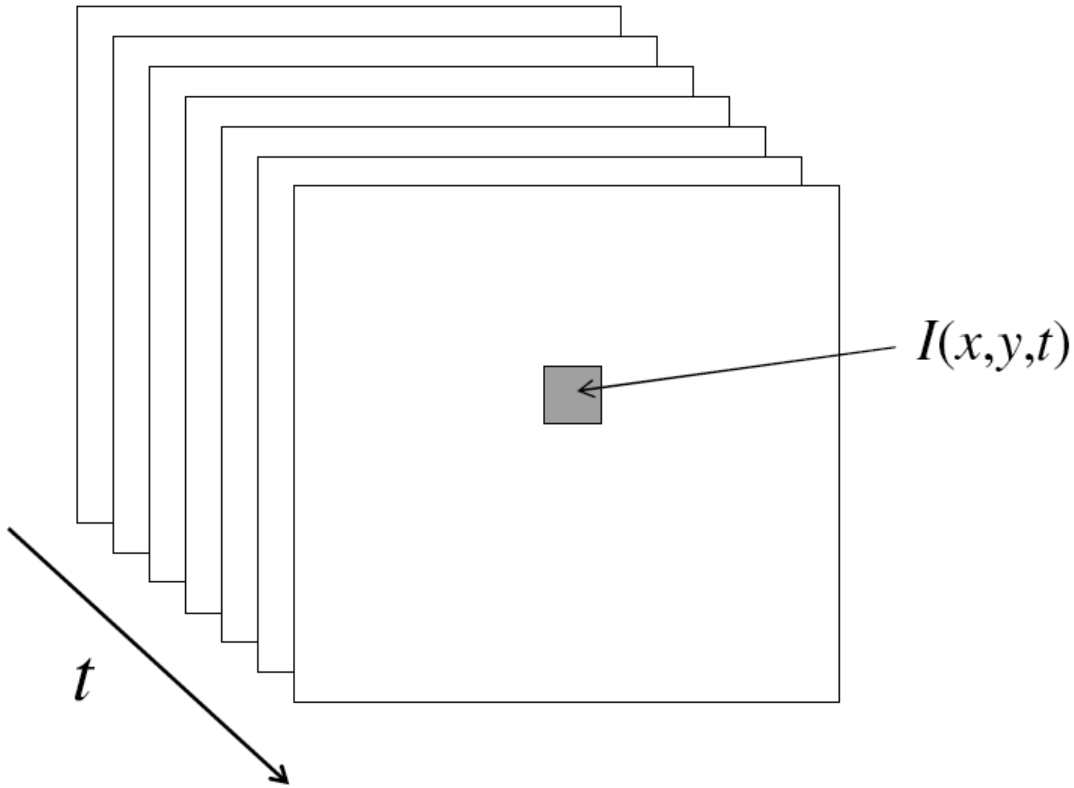
Motion Estimation

Luis F. Teixeira | M.EIC | FEUP

Many slides adapted from S. Seitz, R. Szeliski, M. Pollefeys,
K. Grauman, J. Hays, among others

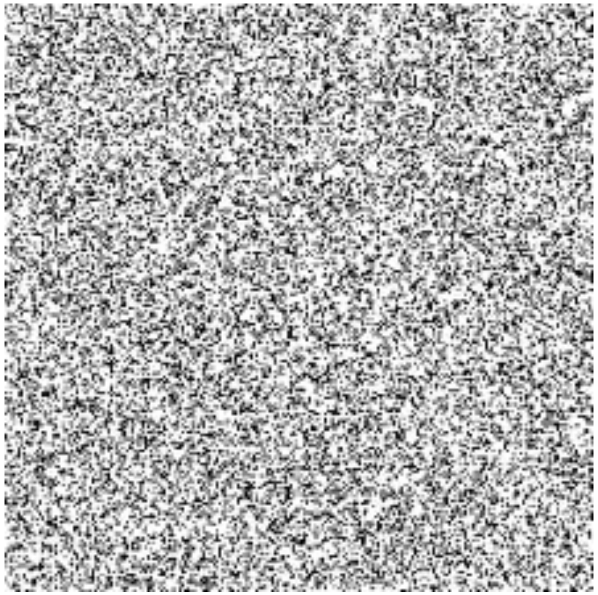
Image Sequences

- A **video** is a **sequence of images** (frames) captured over time
- Now our image data is a function of **space** (x, y) and **time** (t)



Motion and perceptual organization

- Sometimes, motion is the only cue



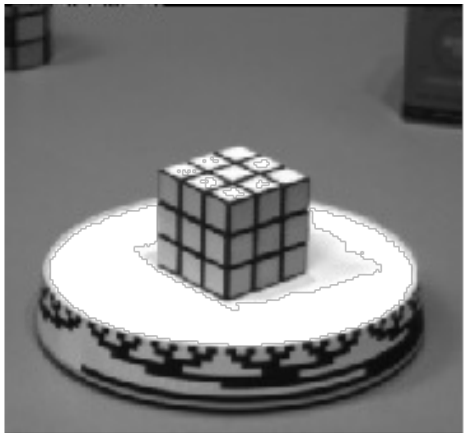
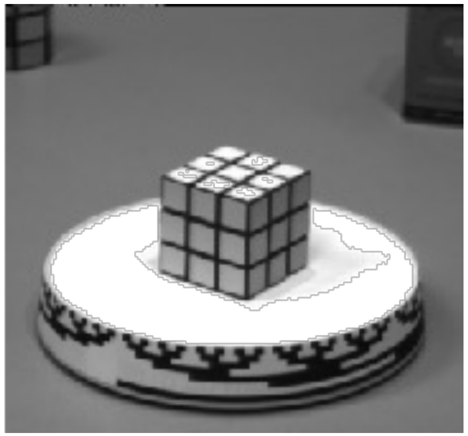
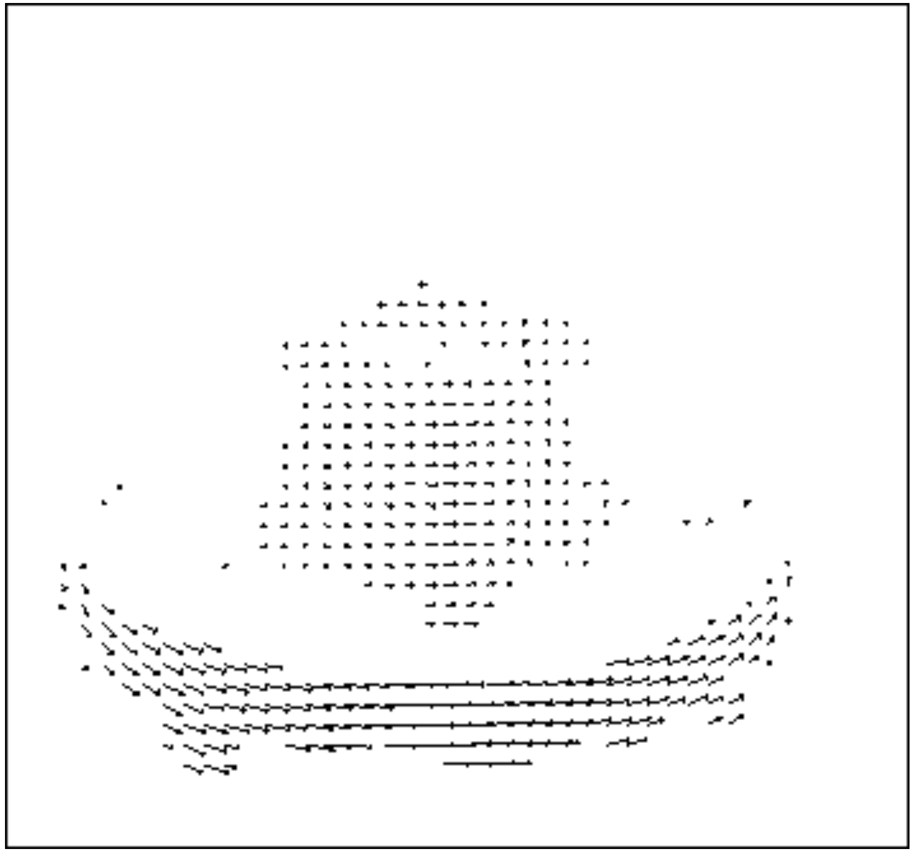
Motion and perceptual organization

- Even “impoverished” motion data can evoke a strong percept

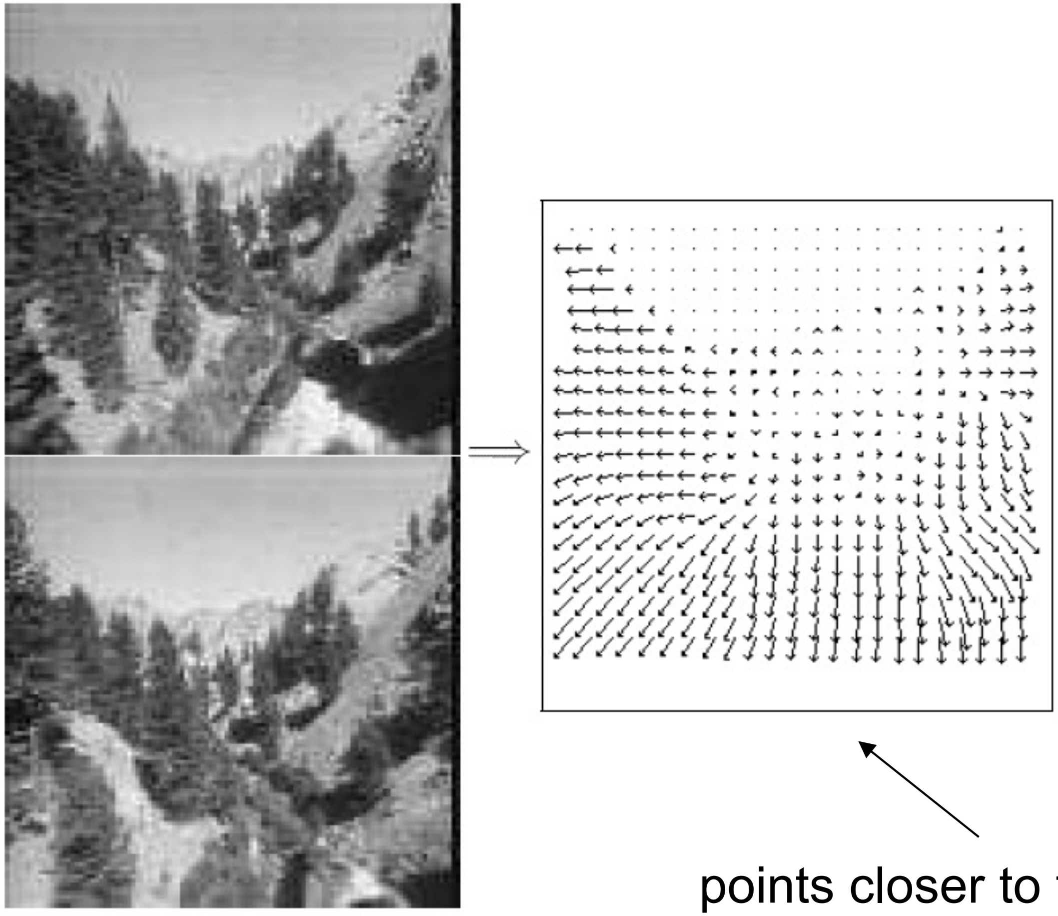
G. Johansson, “Visual Perception of Biological Motion and a Model For Its Analysis", *Perception and Psychophysics 14, 201-211, 1973.*

Motion field

- The motion field is the projection of the 3D scene motion into the image

Motion field + camera motion



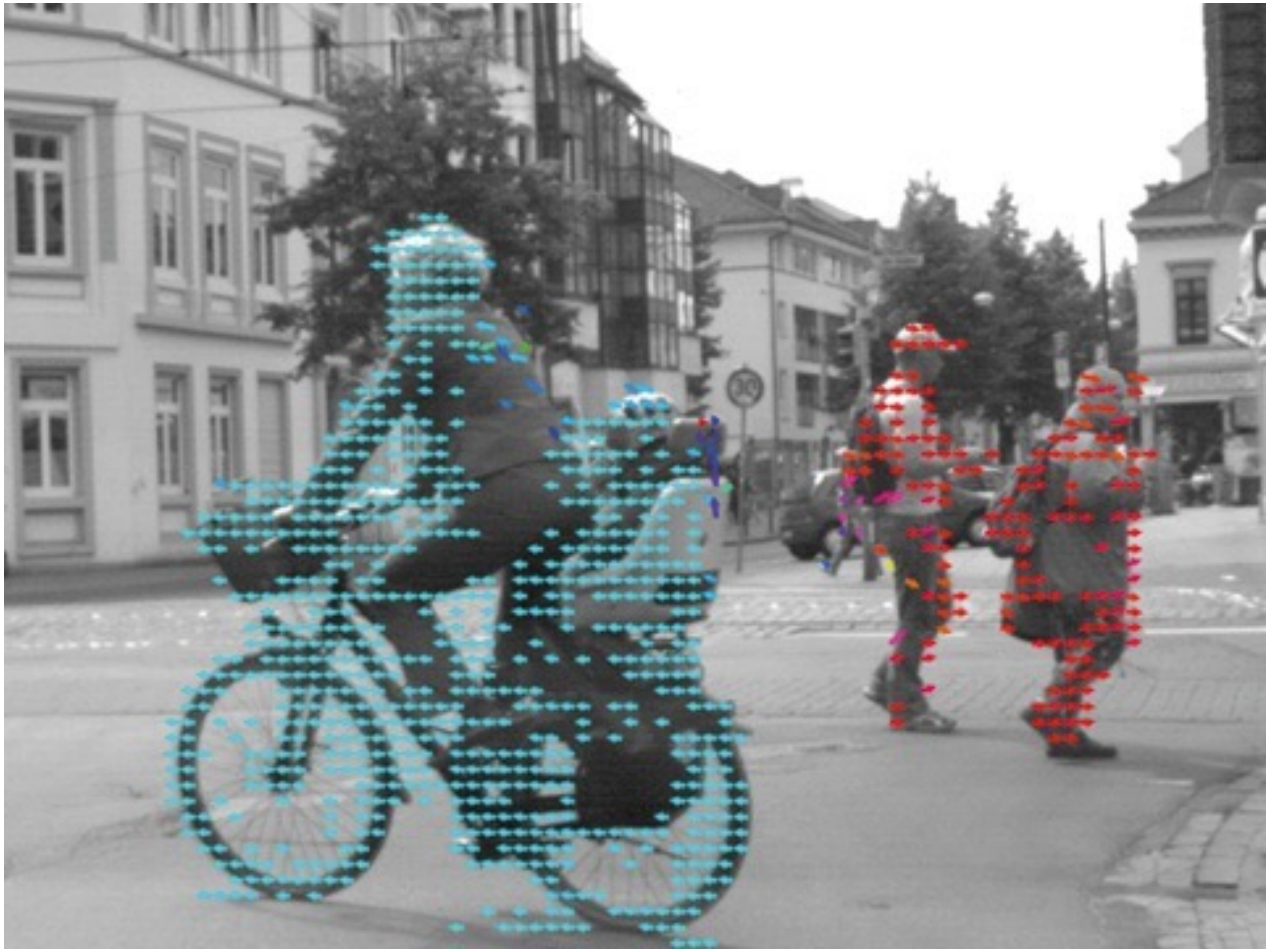
Length of flow vectors inversely proportional to depth Z of 3D point

points closer to the camera move more quickly across the image plane → Motion parallax

Figure from Michael Black, Ph.D. Thesis

Optical flow

Vector field function of the spatio-temporal image
brightness variations



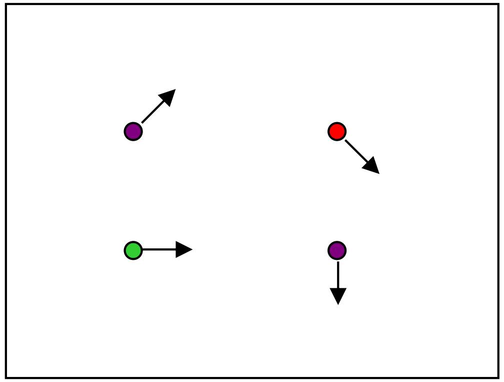
Optical flow

- **Definition**: optical flow is the *apparent* motion of brightness patterns in the image
- **Ideally**, optical flow would be the same as the motion field
- **However**, the apparent motion can be caused by lighting changes without any actual motion

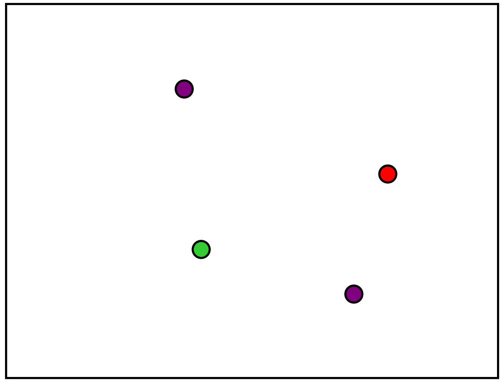
Estimating the Optical flow

- Let’s start by considering each pixel separately
- How do we estimate pixel motion?
  - Feature matching
  - Can we use what we have seen before?
    - Not really… although it works well for matching, detection, recognition
    - Sparse

Feature tracking



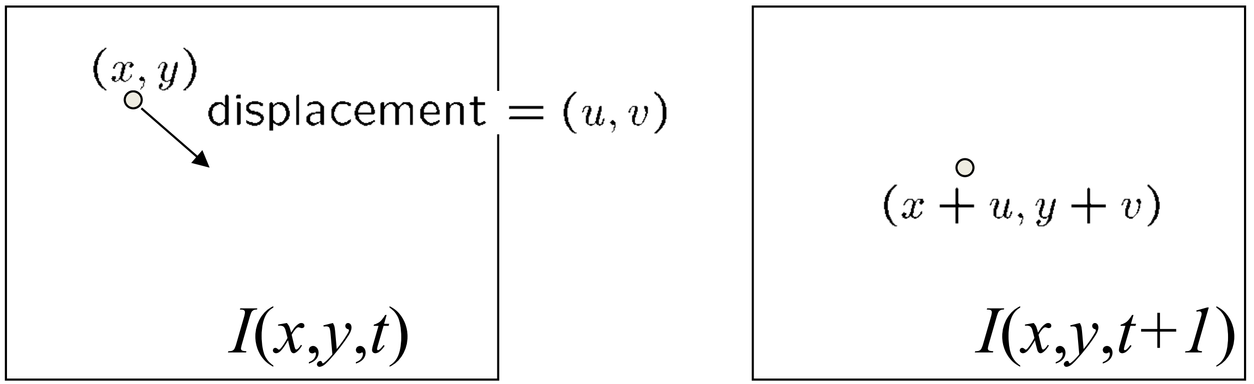
$I(x,y,t)$



$I(x,y,t+1)$

- Given two subsequent frames, estimate the point translation
- Key assumptions of Lucas-Kanade Tracker
  - Brightness constancy: projection of the same point looks the same in every frame
  - Small motion: points do not move very far
  - Spatial coherence: points move like their neighbors

The brightness constancy constraint



Brightness Constancy Equation:

$$I(x, y, t) = I(x+u, y+v, t+1)$$

Small motion (one pixel):

Take first-order Taylor approximation and rearrange expression

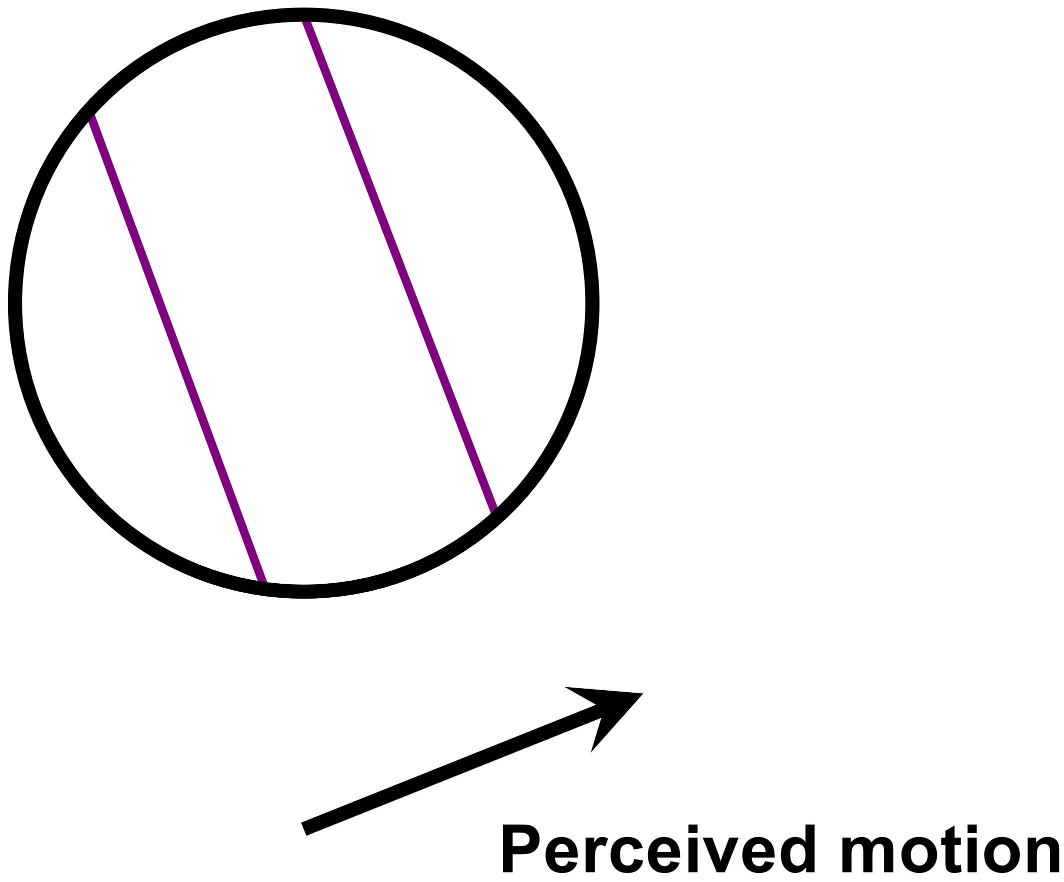
$$I_x u + I_y v + I_t = 0$$

$$\boxed{\nabla \mathrm{I}} \cdot [\mathrm{u} \;\; \mathrm{v}]^{\mathrm{T}} + \boxed{\mathrm{I_t}} = 0$$

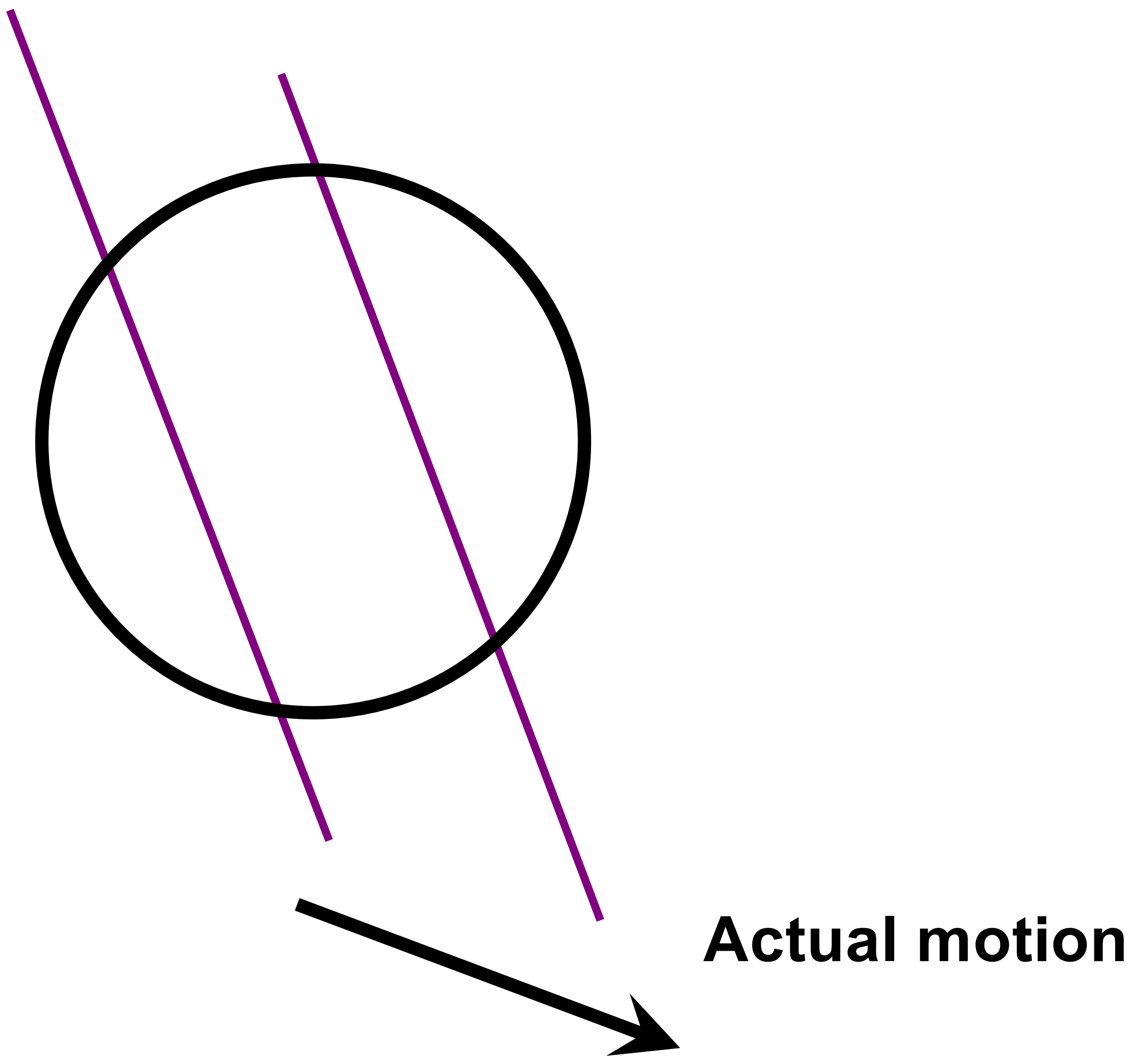
Image derivatives Difference over frames

One equation, two unknowns (u,v)

The aperture problem



The aperture problem



The barber pole illusion

http://en.wikipedia.org/wiki/Barberpole_illusion

The barber pole illusion

Image source: https://en.wikipedia.org/wiki/Barber's_pole

Solving the ambiguity

- How to get more equations for a pixel?
- **Spatial coherence constraint**
- Assume the pixel's neighbors have the same (u,v)
  - If we use a 5x5 window, that gives us 25 equations per pixel

$$\underset{25\times2}{A}\ \underset{2\times1}{d} = \underset{25\times1}{b} \qquad \begin{bmatrix} I_x(\mathbf{p_1}) & I_y(\mathbf{p_1}) \\ I_x(\mathbf{p_2}) & I_y(\mathbf{p_2}) \\ \vdots & \vdots \\ I_x(\mathbf{p_{25}}) & I_y(\mathbf{p_{25}}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p_1}) \\ I_t(\mathbf{p_2}) \\ \vdots \\ I_t(\mathbf{p_{25}}) \end{bmatrix}$$

Overconstrained linear system

Least squares solution

$$\underbrace{\begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix}}_{A^T A} \begin{bmatrix} u \\ v \end{bmatrix} = - \underbrace{\begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix}}_{A^T b}$$

The summations are over all pixels in the K x K window

B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

Conditions for solvability

Optimal (u, v) satisfies Lucas-Kanade equation

$$\begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix}$$

$$A^T A \qquad\qquad\qquad A^T b$$

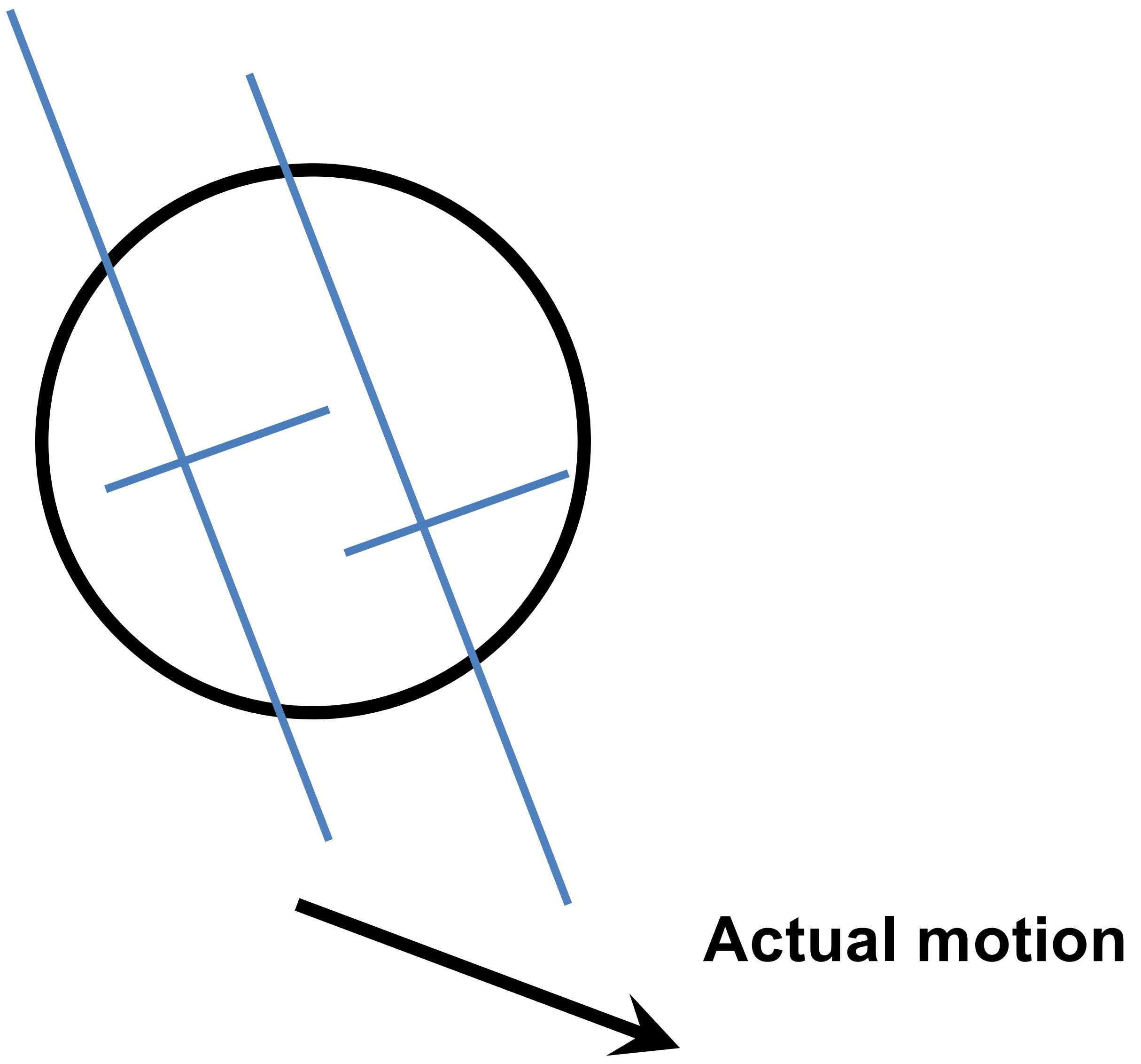
When is this solvable? I.e., what are good points to track?

- $A^T A$ should be invertible
- $A^T A$ should not be too small due to noise
  - eigenvalues λ_1 and λ_2 of $A^T A$ should not be too small
- $A^T A$ should be well-conditioned
  - λ_1 / λ_2 should not be too large (λ_1 = larger eigenvalue)

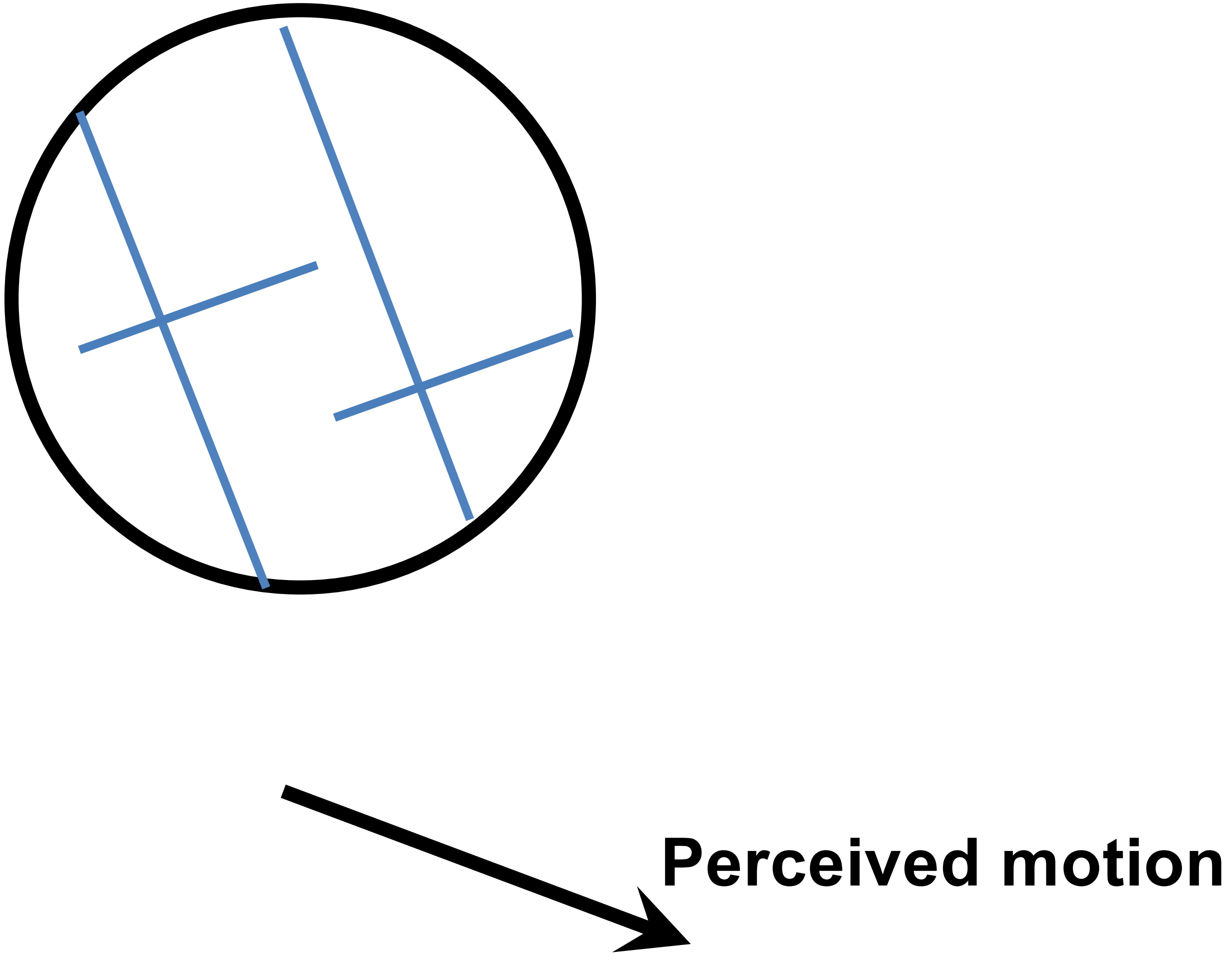
Does this remind you of anything?

Criteria for Harris corner detector

The aperture problem resolved



The aperture problem resolved

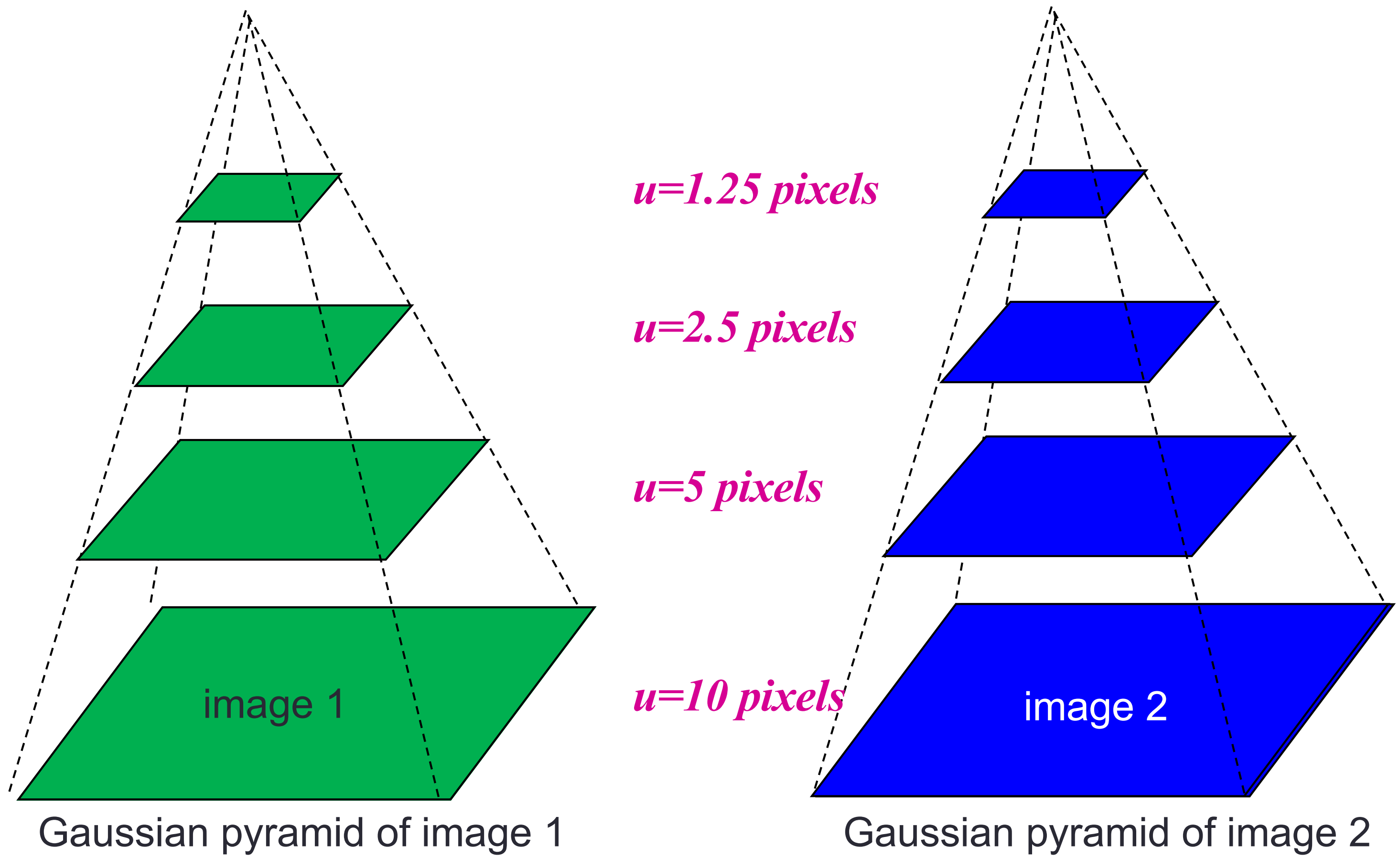


Lucas-Kanade Optical Flow

- Lucas-Kanade feature tracking, but for each pixel
  - Operations can be done one frame at a time, rather than pixel by pixel

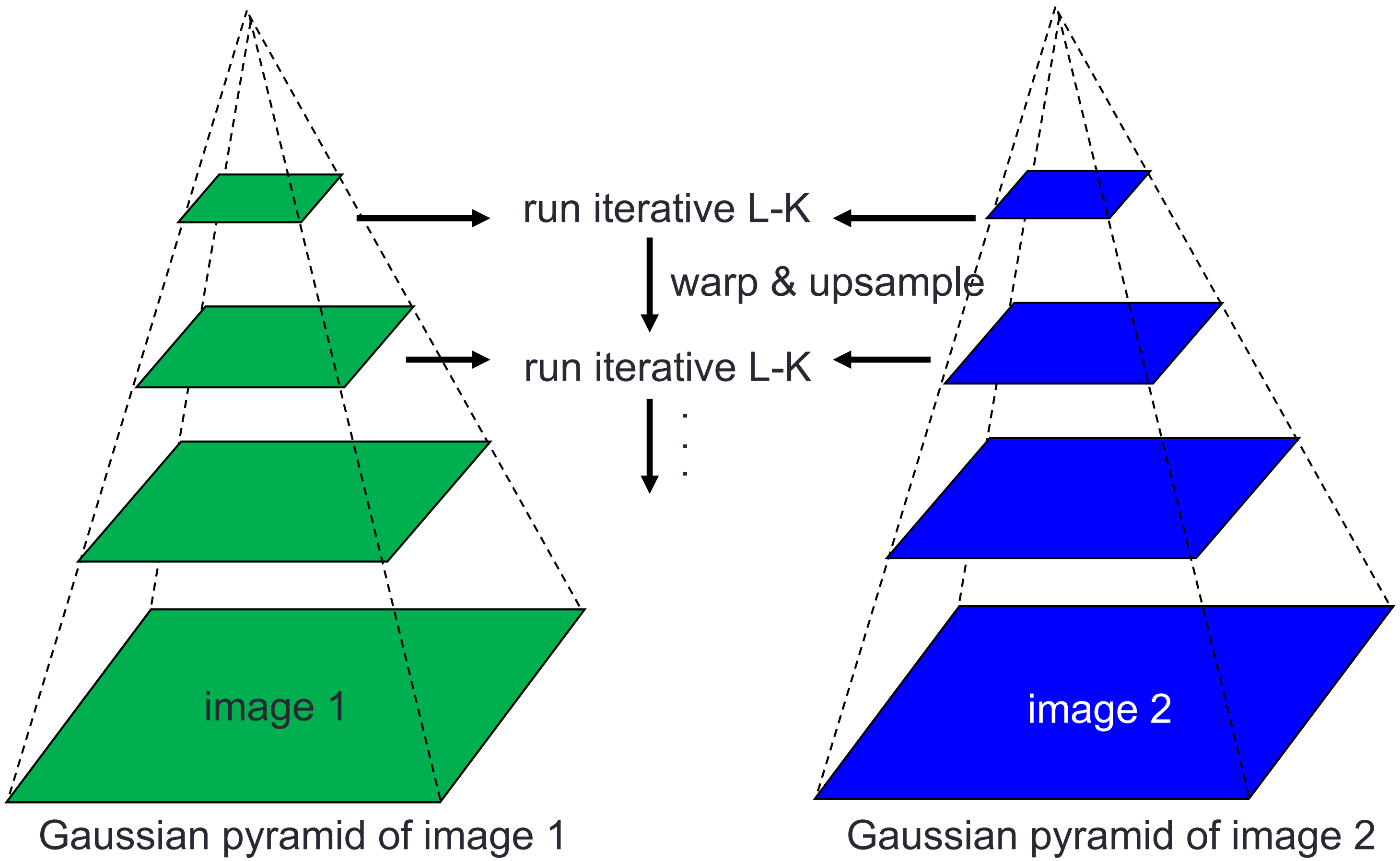
- Revisiting the small motion assumption:
  - What if the motion is large (larger than a pixel)?
    - Not-linear: Iterative refinement – estimate, warp, repeat until convergence
    - Local minima: coarse-to-fine estimation

Coarse-to-fine optical flow estimation

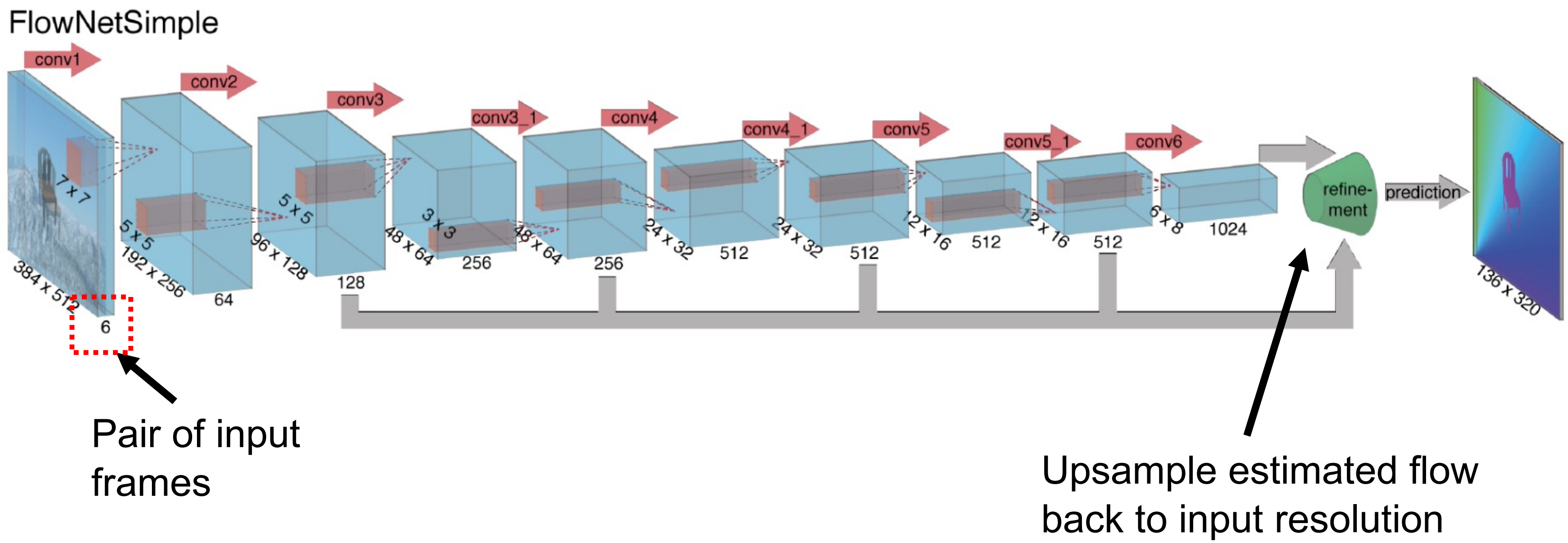


A coarse-to-fine optical flow estimation, using a Gaussian pyramid, helps solve ambiguities when the ”small motion” assumption does not hold

Coarse-to-fine optical flow estimation

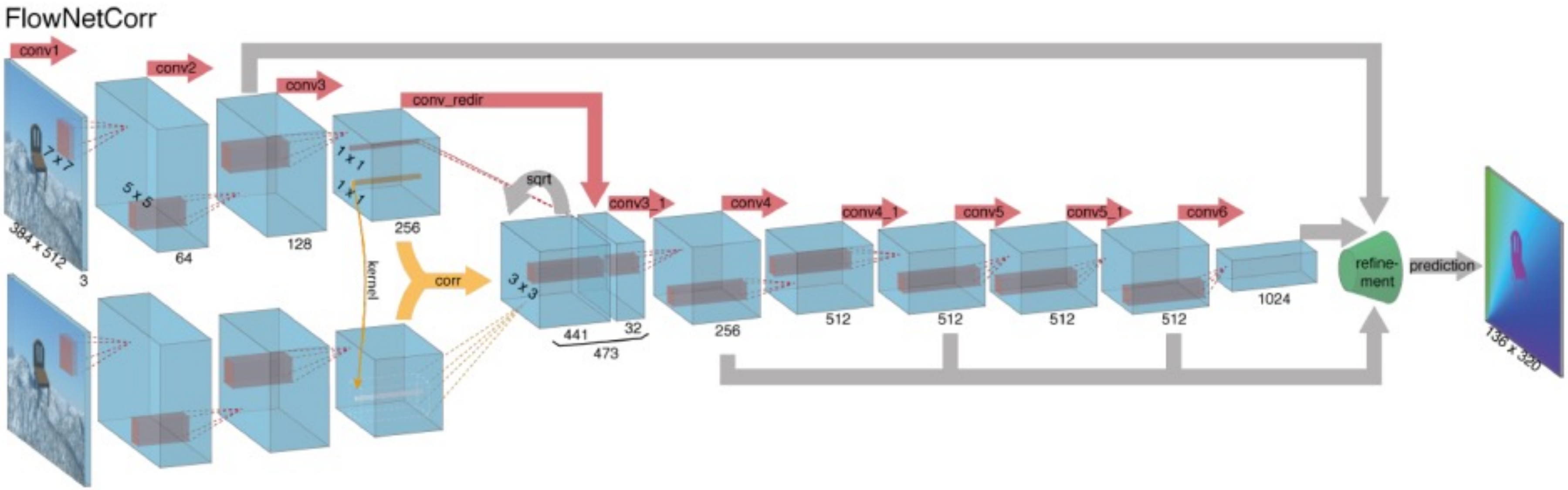


CNN to estimate optical flow



Fischer et al. 2015. https://arxiv.org/abs/1504.06852

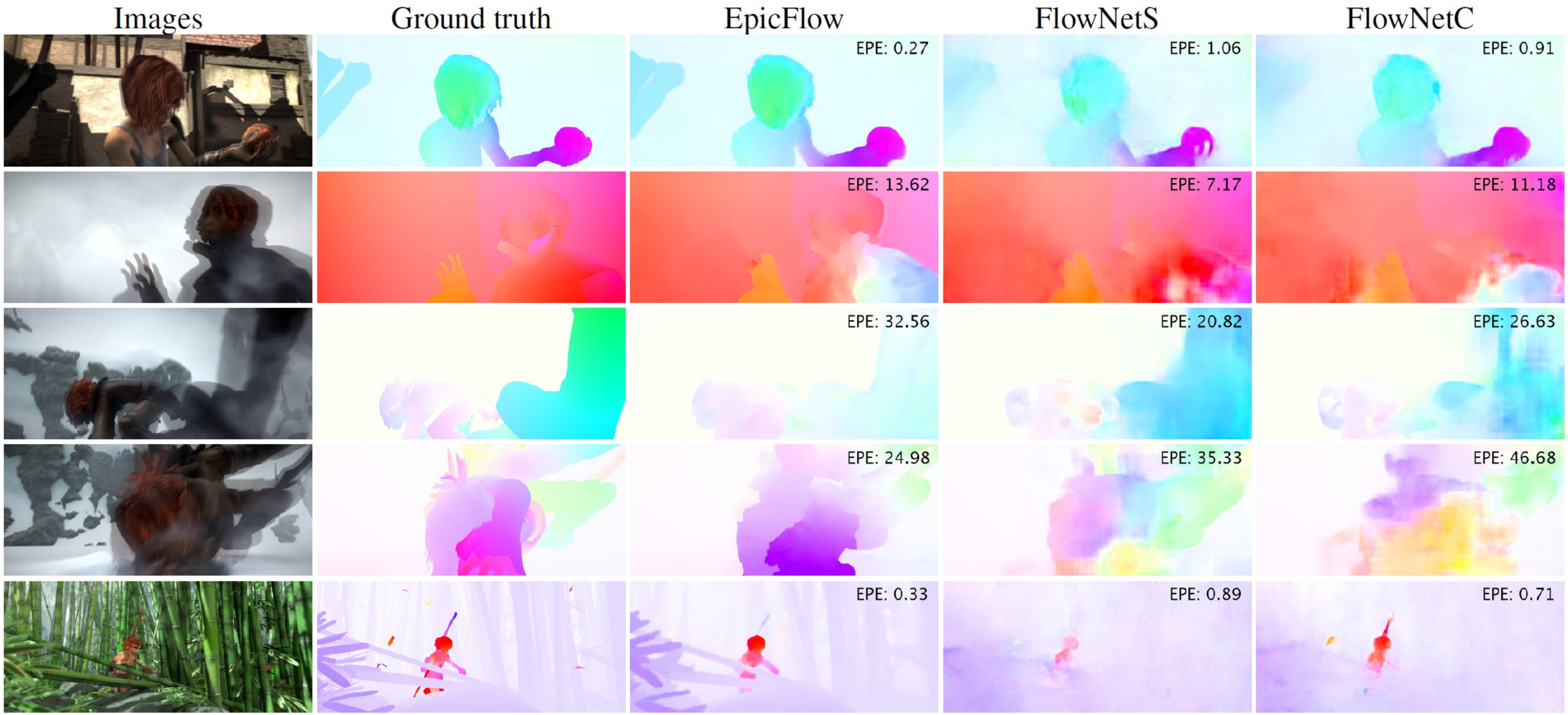
CNN to estimate optical flow



Alternatively, extract features separately and combine later

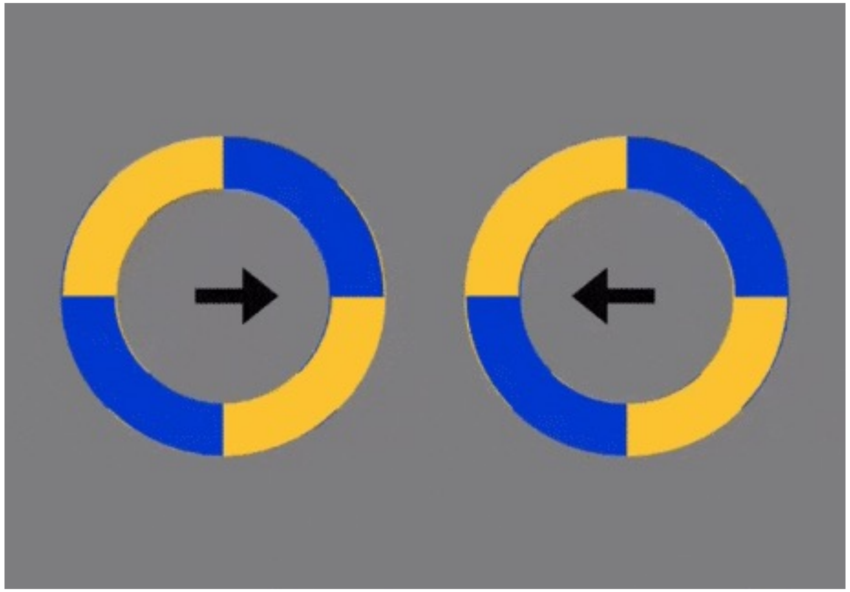
CNN to estimate optical flow

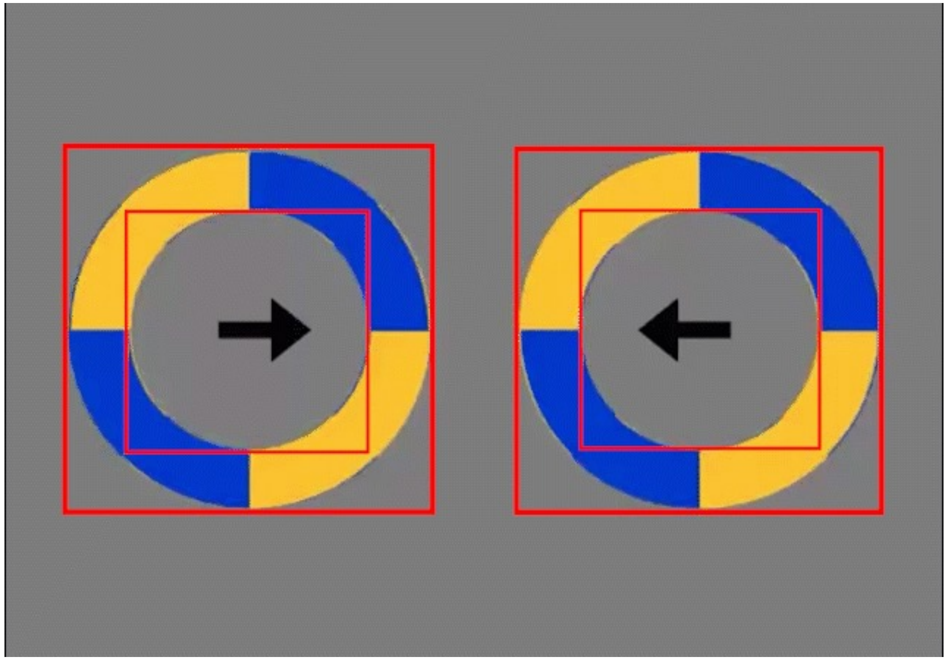
Results on Sintel (open source 3D animated short film)

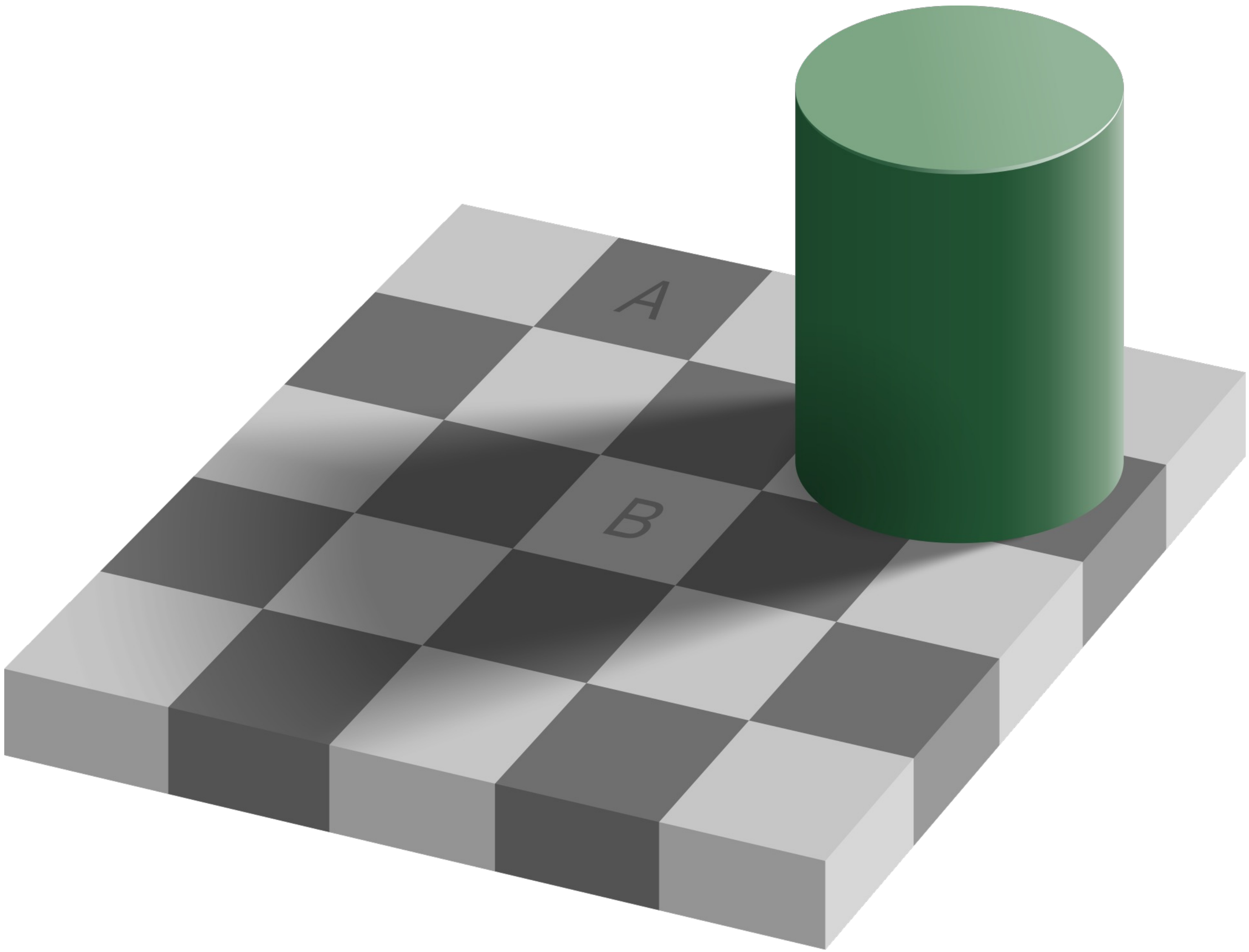


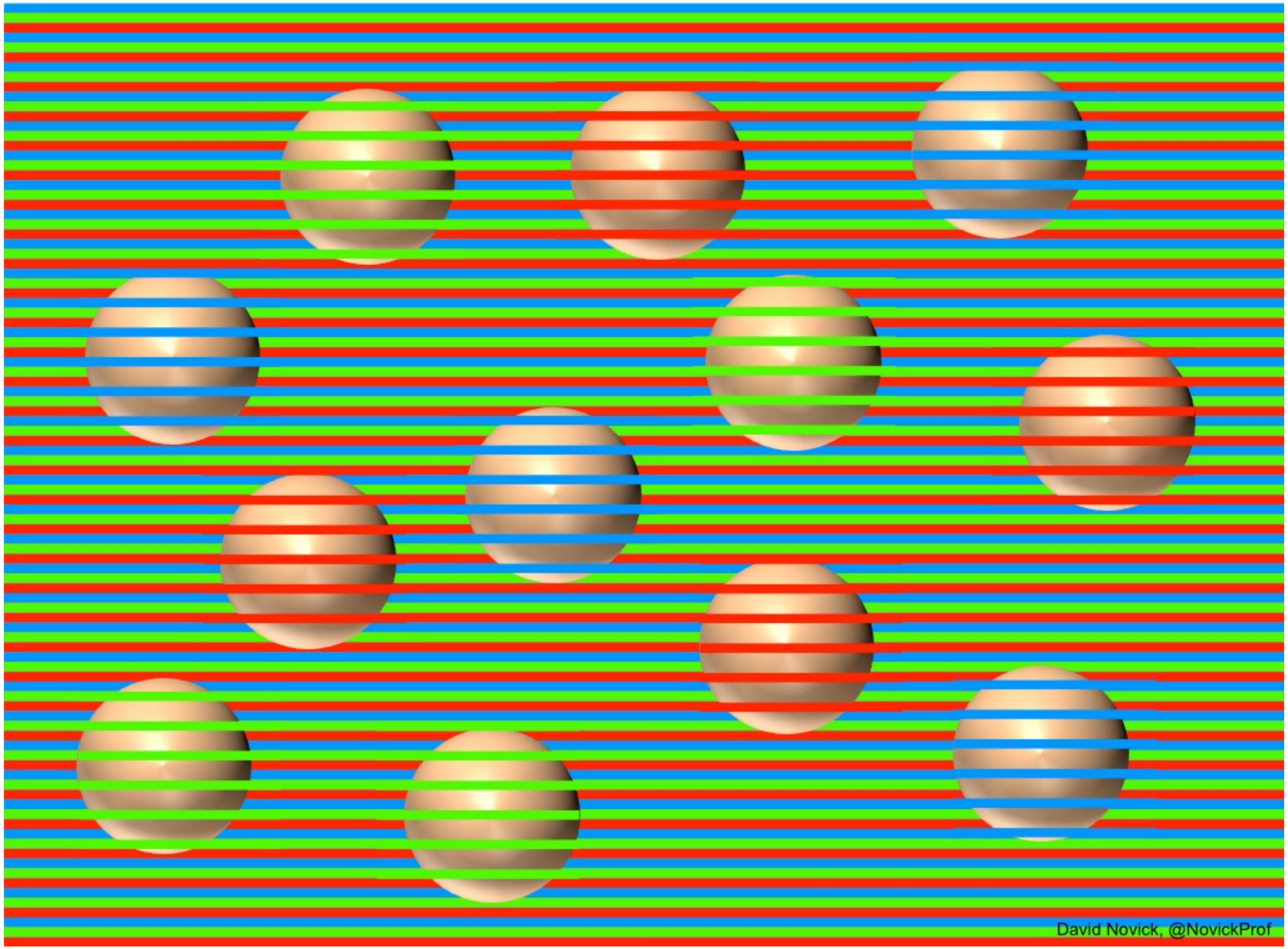
Fischer et al. 2015. https://arxiv.org/abs/1504.06852


And Now for Something
Completely Different






A
B


David Novick, @NovickProf